APPROXIMATION FOR THE DISTRIBUTION OF EXTREMES OF ONE DEPENDENT STATIONARY SEQUENCES OF RANDOM VARIABLES

AMĂRIOAREI ALEXANDRU

Abstract. In this paper we improve some existing results concerning the approximation of the distribution of extremes of a 1-dependent and stationary sequence of random variables. We enlarge the range of applicability and improve the approximation error. An application to the study of the distribution of scan statistics generated by Bernoulli trials is given.

1. Introduction

The starting point of this paper is a series of results of Haiman [1999], concerning the extreme of a 1-dependent and stationary sequence of random variables. Let $(X_n)_{n\geq 1}$ be a sequence of strictly stationary 1-dependent random variables (for any $t \geq 1$ we have $\sigma(X_1,\dots,X_t)$ and $\sigma(X_{t+2},\dots)$ are independent) with marginal distribution function $F(x) = \mathbb{P}(X_1 \leq x)$. Let x such that

$$\inf\{u|F(u) > 0\} < x < \sup\{u|F(u) < 1\}.$$

Define the sequences

$$\begin{aligned} p_n &= p_n(x) = \mathbb{P}(\min\{X_1, X_2, \dots, X_n\} > x), \quad n \geq 1, \; p_0 = 1, && (1.1)\\ q_n &= q_n(x) = \mathbb{P}(\max\{X_1, X_2, \dots, X_n\} \leq x), \quad n \geq 1 && (1.2) \end{aligned}$$

and the series

$$C(z) = C_x(z) = 1 + \sum_{k=1}^{\infty} (-1)^k p_{k-1} z^k. \tag{1.3}$$

In Haiman [1999], the author proved the following results:

Theorem 1.1. *For x such that $0 < p_1(x) \leq 0.025$, $C_x(z)$ has a unique zero $\lambda(x)$, of order of multiplicity 1, inside the interval $(1, 1+2p_1)$, such that*

$$|\lambda - (1 + p_1 - p_2 + p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2)| \leq 87p_1^3. \tag{1.4}$$

Theorem 1.2. *We have*

$$q_1 = 1 - p_1, \qquad q_2 = 1 - 2p_1 + p_2, \qquad q_3 = 1 - 3p_1 + 2p_2 + p_1^2 - p_3$$

and for $n > 3$ if $p_1 \leq 0.025$,

$$|q_n\lambda^n - (1 - p_2 + 2p_3 - 3p_4 + p_1^2 + 6p_2^2 - 6p_1p_2)| \leq 561p_1^3. \tag{1.5}$$

2000 *Mathematics Subject Classification.* 62E17.
Key words and phrases. extremes, 1-dependent stationary sequences, scan statistics.

These results were successfully applied in a series of applications: the distribution of the maximum of the increments of the Wiener process (Haiman [1999]), extremes of Markov sequences (Haiman et al. [1995]), the distribution of scan statistics, both in one dimensional (see Haiman [2000, 2007]) and two dimensional case (see Haiman and Preda [2002, 2006]).
Following the same lines of proofs as in Haiman [1999], we improve Theorem 1.1 and Theorem 1.2 by enlarging the range of applicability and providing sharper error bounds. The main results are presented in Section 2. In Section 3, we present an application to the study of the distribution of one dimensional discrete scan statistics emphasizing the difference between the new and the old results. Proofs are presented in Section 4.

2. Main results

The following theorem gives a parametric form of the Theorem 1.1 improving both the range of $p_1(x)$, from 0.025 to 0.1, and the error coefficient:

Theorem 2.1. *For x such that $0 < p_1(x) \leq \alpha \leq 0.1$, $C_x(z)$ has an unique zero $\lambda(x)$, of order of multiplicity 1, inside an interval of the form $(1, 1 + lp_1)$, such that*

$$|\lambda - (1 + p_1 - p_2 + p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2)| \leq K(\alpha)p_1^3, \tag{2.1}$$

where $l = l(\alpha) > t_2^3(\alpha)$, $t_2(\alpha)$ is the second root in magnitude of the equation $\alpha t^3 - t + 1 = 0$ and $K(\alpha)$ is given by

$$K(\alpha) = \frac{\frac{11-3\alpha}{(1-\alpha)^2} + 2l(1+3\alpha)\frac{2+3l\alpha-\alpha(2-l\alpha)(1+l\alpha)^2}{[1-\alpha(1+l\alpha)^2]^3}}{1 - \frac{2\alpha(1+l\alpha)}{[1-\alpha(1+l\alpha)^2]^2}}. \tag{2.2}$$

Using the properties of the 1-dependent sequence, an immediate consequence of Theorem 2.1 is the following:

Corollary 2.2. *Let λ be defined as in Theorem 2.1, then*

$$|\lambda - (1 + p_1 - p_2 + 2(p_1 - p_2)^2)| \leq (1 + \alpha K(\alpha))p_1^2. \tag{2.3}$$

To get a better grasp of the bounds in Theorem 2.1 and Corollary 2.2, we present, for selected values of α, the values taken by the coefficients in Eq.(2.1) and Eq.(2.3):

α	l	$K(\alpha)$	$1 + \alpha K(\alpha)$
0.100	1.5347	38.6302	4.8630
0.050	1.1893	21.2853	2.0642
0.025	**1.0835**	**17.5663**	**1.4391**
0.010	1.0313	15.9265	1.1592

Table 1. Selected values for the error coefficients in Theorem 2.1 and Corollary 2.2

Notice that for the value considered in Haiman [1999], i.e. $\alpha = 0.025$, our corresponding value for the error coefficient in Eq.(2.1) is almost five times smaller than in Eq.(1.4). The following result improves Theorem 1.2:

Theorem 2.3. *Lets suppose that x is such that $0 < p_1(x) \leq \alpha \leq 0.1$ and define $\eta = 1 + l\alpha$ with $l = l(\alpha) > t_2^3(\alpha)$ and $t_2(\alpha)$ the second root in magnitude of the equation $\alpha t^3 - t + 1 = 0$. If $\lambda = \lambda(x)$ is the zero obtained in Theorem 2.1, then the following relation holds*

$$|q_n\lambda^n - (1 - p_2 + 2p_3 - 3p_4 + p_1^2 + 6p_2^2 - 6p_1p_2)| \leq \Gamma(\alpha)p_1^3, \tag{2.4}$$

where $\Gamma(\alpha) = L(\alpha) + E(\alpha)$, $K(\alpha)$ is given by Eq.(2.2) and

$$\begin{aligned} L(\alpha) &= 3K(\alpha)(1+\alpha+3\alpha^2)[1+\alpha+3\alpha^2+K(\alpha)\alpha^3]+\alpha^6K^3(\alpha) \\ &\quad +9\alpha(4+3\alpha+3\alpha^2)+55 \end{aligned} \tag{2.5}$$

$$E(\alpha) = 0.1 + \frac{\eta^5\left[1+(1-2\alpha)\eta\right]^4\left[1+\alpha(\eta-2)\right]\left[1+\eta+(1-3\alpha)\eta^2\right]}{2(1-\alpha\eta^2)^4\left[(1-\alpha\eta^2)^2-\alpha\eta^2(1+\eta-2\alpha\eta)^2\right]} \tag{2.6}$$

The next corollary is an immediate consequence of Theorem 2.3:

Corollary 2.4. *In the conditions of Theorem 2.3 we have*

$$|q_n\lambda^n - (1 - p_2)| \leq (3 + \alpha\Gamma(\alpha))p_1^2. \tag{2.7}$$

The error coefficient in Eq.(2.4) is smaller in comparison with the corresponding one from Eq.(1.5), as the following table can show:

α	$\Gamma(\alpha)$	$3+\alpha\Gamma(\alpha)$
0.100	480.696	51.0696
0.050	180.532	12.0266
0.025	**145.202**	**6.6300**
0.010	131.438	4.3143

TABLE 2. Selected values for the error coefficients in Theorem 2.3 and Corollary 2.4

Combining the results obtained in Theorem 2.1 and Theorem 2.3 we get the following approximation:

Theorem 2.5. *Let x such that $q_1(x) \geq 1 - \alpha \geq 0.9$. If $\Gamma(\alpha)$ and $K(\alpha)$ are the same as in Theorem 2.3, then*

$$\left| q_n - \frac{6(q_1-q_2)^2 + 4q_3 - 3q_4}{(1 + q_1 - q_2 + q_3 - q_4 + 2q_1^2 + 3q_2^2 - 5q_1q_2)^n} \right| \leq \Delta_1(1-q_1)^3 \tag{2.8}$$

with

$$\Delta_1 = \Delta_1(\alpha, n) = \Gamma(\alpha) + nK(\alpha). \tag{2.9}$$

In the same fashion combining the results from Corollary 2.2 and Corollary 2.4 we get

Theorem 2.6. *If x is such that $q_1(x) \geq 1 - \alpha \geq 0.9$, then*

$$\left| q_n - \frac{2q_1 - q_2}{[1 + q_1 - q_2 + 2(q_1-q_2)^2]^n} \right| \leq \Delta_2(1-q_1)^2 \tag{2.10}$$

with

$$\Delta_2 = \Delta_2(\alpha, n, q_1) = 3 + \Gamma(\alpha)(1-q_1) + n\left[1 + K(\alpha)(1-q_1)\right], \tag{2.11}$$

and where $\Gamma(\alpha)$ and $K(\alpha)$ are the same as in Theorem 2.3.

3. Application to the distribution of scan statistics

In many applications the decision makers have to determine if a certain accumulation of events is *normal* or not, where by *normal* we mean that it can be explained by an underlying probability model defined by a null hypothesis of randomness. One way of determining if such a cluster of observations is exceptionally or not is by using the scan statistics. For a comprehensive methodological treatment and a rich source of applications of this test statistic, one may study the books of Glaz, Naus and Wallenstein [2001] and more recently the one of Glaz, Pozdnyakov and Wallenstein [2009].

Let $Y_1, Y_2, \dots, Y_N$ be a sequence of independent and identically distributed (i.i.d.) random variables and m, $1 \leq m \leq N$, be a fixed positive integer. If we consider the random variables

$$Z_t = \sum_{i=t}^{t+m-1} Y_i, \quad 1 \leq t \leq N-m+1 \tag{3.1}$$

then the one dimensional discrete scan statistic is defined by

$$S_m(N) = \max_{1 \leq t \leq N-m+1} Z_t. \tag{3.2}$$

To get an intuitive meaning of the above definition, if Y_i are integer valued random variables then we can interpret them as the number of observed events at the time i. The scan statistics is then viewed as the maximum number of events observed in any contiguous period of length m within the interval $\{1, 2, \dots, N\}$. Since exact formulas for the distribution of S_m exist only in a small number of situations (see for example Glaz, Naus and Wallenstein [2001], Chapter 13), various approximation methods and bounds have been proposed. In what follows we will use the approximation method developed in Haiman [2007] but with the help of the results obtained in Section 2. The method is based on the important observation that, in the i.i.d. case, the discrete scan statistic can be expressed as an extreme of a 1-dependent stationary sequence. It is easy to see that the random variables

$$W_k = \max_{(k-1)m+1 \leq s \leq km+1} Z_s, \quad k = 1, 2, \dots \tag{3.3}$$

form a 1-dependent stationary sequence and that the following relation holds,

$$S_m\left(Lm\right) = \max\left(W_1, W_2, \dots, W_{L-1}\right). \tag{3.4}$$

Observe that for general N, one can consider $L = [N/m]$ and then apply the inequality

$$\mathbb{P}\left(S_m[(L+1)m] \leq n\right) \leq \mathbb{P}\left(S_m(N) \leq n\right) \leq \mathbb{P}\left(S_m(Lm) \leq n\right). \tag{3.5}$$

When $\mathbb{P}(W_1 > n) \leq 0.1$, we can apply the results from Theorem 2.6 to the sequence $W_1, W_2, \dots, W_{L-1}$ to obtain the following approximation for the distribution of the scan statistic S_m:

$$\mathbb{P}\left(S_m(Lm) \leq n\right) \approx \frac{2q_1 - q_2}{\left[1 + q_1 - q_2 + 2(q_1 - q_2)^2\right]^{(L-1)}}, \tag{3.6}$$

with an error bound of about

$$E = \left\{3 + \Gamma(\alpha)(1-q_1) + (L-1)\left[1 + K(\alpha)(1-q_1)\right]\right\}(1-q_1)^2, \tag{3.7}$$

where $\alpha = \mathbb{P}(W_1 > n)$ and q_1, q_2 are defined by Eq.(1.2) as

$$q_1 = \mathbb{P}(W_1 \leq n) = \mathbb{P}(S_m(2m) \leq n), \quad (3.8)$$
$$q_2 = \mathbb{P}(W_1 \leq n, W_2 \leq n) = \mathbb{P}(S_m(3m) \leq n). \quad (3.9)$$

We should mention that in Haiman [1999, Theorem 4], the author obtained the following formula for the approximation error

$$EH = \left\{9 + 561(1-q_1) + 3.3(L-1)\left[1 + 4.7(L-1)(1-q_1)^2\right]\right\}(1-q_1)^2. \quad (3.10)$$

Notice that if q_1, q_2, q_3, q_4 are known then we can apply Theorem 2.5 to get a better approximation for the distribution of S_m, even though in most applications Eq.(3.6) will suffice.

Next we will restrict ourselves to the case of Bernoulli $0-1$ process, that is Y_i's are Bernoulli trials with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$. In this particular framework, Naus (see for example Glaz, Naus and Wallenstein [2001], Theorem 13.1) provided exact formulas for $\mathbb{P}(S_m(N) \leq n)$ when $N = 2m$ and $N = 3m$, namely for q_1 and q_2 from Eq.(3.8) and Eq.(3.9). The following tables illustrates a compared study between the error formula given by Eq.(3.7) and the corresponding error bound used by Haiman [2007] along with the exact and approximated value for the distribution of the scan statistic:

n	q_1 Eq.(3.8)	q_2 Eq.(3.9)	$Approx$ Eq.(3.6)	$Exact$	EH Eq.(3.10)	E Eq.(3.7)
2	0.97131	0.95181	0.82715	0.82582	–	0.01712
3	0.99716	0.99500	0.98001	0.98000	0.00032	0.00010
4	0.99982	0.99967	0.99865	0.99865	1×10^{-6}	3×10^{-7}
5	0.99999	0.99998	0.99994	0.99994	2×10^{-9}	6×10^{-10}
6	1.	1.	0.99999	0.99999	1×10^{-12}	4×10^{-13}
7	1.	1.	1.	1.	3×10^{-16}	9×10^{-17}

TABLE 3. Distribution of the scan statistic $\mathbb{P}(S_m(Lm) \leq n)$ for $m = 9$, $p = 0.05$, $L = 10$

n	q_1 Eq.(3.8)	q_2 Eq.(3.9)	$Approx$ Eq.(3.6)	$Exact$	EH Eq.(3.10)	E Eq.(3.7)
1	0.96860	0.94910	0.74617	0.74353	–	0.02927
2	0.99813	0.99677	0.98061	0.98060	0.00019	0.00006
3	0.99993	0.99987	0.99922	0.99922	2×10^{-7}	8×10^{-8}
4	0.99999	0.99999	0.99998	0.99998	1×10^{-10}	4×10^{-11}
5	1.	1.	1.	1.	4×10^{-14}	1×10^{-14}

TABLE 4. Distribution of the scan statistic $\mathbb{P}(S_m(Lm) \leq n)$ for $m = 10$, $p = 0.0165$, $L = 15$

The exact values for the distribution of the scan statistics presented in the tables 3 and 4 (column "Exact") are computed using the Markov chain embedding technique described in Fu [2001].

4. Proofs of the results

4.1. **Proof of Theorem 2.1.** The proof will follow closely that of Haiman [1999]. Using the stationarity and 1-dependence of the sequence we have

$$\begin{aligned} p_n = \mathbb{P}(X_1 > x, X_2 > x, \ldots, X_n > x) &\leq \mathbb{P}(X_1 > x, X_3 > x, \ldots, X_n > x) \\ &= \mathbb{P}(X_1 > x)\mathbb{P}(X_3 > x, \ldots, X_n > x) \\ &= p_1 p_{n-2}, \end{aligned}$$

which give the basic inequality

$$p_n \leq p_1^{[\frac{n+1}{2}]}. \tag{4.1}$$

To show that $C(z)$ has a zero in the interval $(1, 1 + lp_1)$ it is enough to prove that $C(1) > 0$ and $C(1 + lp_1) < 0$. It easy to see that $C(1) > 0$, since

$$C(1) = 1 + \sum_{k=1}^{\infty} (-1)^k p_{k-1} = \underbrace{(p_1 - p_2)}_{\geq 0} + \underbrace{(p_3 - p_4)}_{\geq 0} + \cdots \geq 0 \tag{4.2}$$

For $C(1 + lp_1)$ we have:

$$\begin{aligned} C(1 + lp_1) &= 1 + \sum_{k=1}^{\infty} (-1)^k p_{k-1} (1 + lp_1)^k \\ &= -lp_1 + \sum_{k=1}^{\infty} (1 + lp_1)^{2k} \underbrace{[p_{2k-1} - p_{2k}(1 + lp_1)]}_{\leq p_{2k-1} \leq p_1^k} \\ &\leq -lp_1 + \sum_{k=1}^{\infty} [(1 + lp_1)^2 p_1]^k. \end{aligned} \tag{4.3}$$

It is easy to see that if $(1 + lp_1)^3 < l$ then $p_1(1 + lp_1)^2 < \frac{lp_1}{1+lp_1} < 1$, so series in Eq.(4.3) is convergent and $C(1 + lp_1) < 0$. From the definition of l and the relation $1 < t_2(\alpha) < t_2(\alpha) + \varepsilon < \frac{1}{\sqrt{3\alpha}}$, we obtain $(t_2 = t_2(\alpha))$

$$\begin{aligned} t_2^2 + t_2 \sqrt[3]{t_2^3 + \varepsilon} + \sqrt[3]{(t_2^3 + \varepsilon)^2} &\leq t_2^2 + t_2(t_2 + \varepsilon) + (t_2 + \varepsilon)^2 \\ &< \frac{1}{3\alpha} + \frac{1}{3\alpha} + \frac{1}{3\alpha} = \frac{1}{\alpha} \end{aligned}$$

which imply that $t_2 - \sqrt[3]{t_2^3 + \varepsilon} + \alpha\varepsilon < 0$. Combining this last relation with the fact that t_2 is a root of the equation $\alpha t^3 - t + 1 = 0$, we have that $(1 + lp_1)^3 < l$. For showing that the zero is unique we will prove that $C'(z)$ is strictly decreasing on the interval $(1, 1 + lp_1)$, i.e. $C'(z) < 0$. Using Lagrange theorem on $[1, 1 + a]$ we get for $\theta \in (0, 1)$

$$C'(1 + a) = C'(1) + aC''(1 + \theta a) \tag{4.4}$$

We will approximate both $C'(1)$ and $C''(1 + \theta a)$ as follows:

$$C'(1) = -1 + 2p_1 - 3p_2 + R \tag{4.5}$$

where

$$R = \sum_{k=2}^{\infty} 2k\,(p_{2k-1} - p_{2k}) - \sum_{k=2}^{\infty} p_{2k}. \tag{4.6}$$

To approximate R notice that $p_{2k-1} - p_{2k} \geq 0$, which implies

$$-\sum_{k=2}^{\infty} p_{2k} \leq R \leq \sum_{k=2}^{\infty} 2k\left(p_{2k-1} - p_{2k}\right)$$

and using Eq.(4.1) we get

$$-\frac{p_1^2}{1-p_1} = -\sum_{k=2}^{\infty} p_1^k \leq R \leq 2p_1 \sum_{k=2}^{\infty} k p_1^{k-1} = 2p_1^2\left[\frac{1}{(1-p_1)^2} + \frac{1}{1-p_1}\right] \tag{4.7}$$

so that

$$|R| \leq 2p_1^2\left[\frac{1}{(1-p_1)^2} + \frac{1}{1-p_1}\right]. \tag{4.8}$$

For $C''(z)$ we have:

$$C''(z) = 2p_1 - 2\cdot 3p_2 z + 3\cdot 4p_3 z^2 - 4\cdot 5p_4 z^3 + 5\cdot 6p_5 z^4 - \ldots \tag{4.9}$$

Using $z \in (1, 1+lp_1)$ and Eq.(4.1) we have

$$\begin{aligned} C''(z) &\leq p_1 \sum_{k=0}^{\infty} (2k+1)(2k+2)(z\sqrt{p_1})^{2k} \\ &\leq 2p_1 \frac{1+3p_1(1+lp_1)^2}{\left[1-p_1(1+lp_1)^2\right]^3}. \end{aligned} \tag{4.10}$$

Since $z \leq 1+lp_1$, Eq.(4.9) and Eq.(4.1) implies

$$\begin{aligned} C''(z) &\geq -lp_1 \sum_{k=0}^{\infty} (2k+1)(2k+2)p_{2k+2}z^{2k} - 4\sum_{k=0}^{\infty} (k+1)p_{2k+2}z^{2k+1} \\ &\geq -lp_1^2 \sum_{k=0}^{\infty} (2k+1)(2k+2)(z\sqrt{p_1})^{2k} - 4p_1 z \sum_{k=0}^{\infty} (k+1)(p_1 z^2)^k \\ &\geq -2lp_1^2 \frac{1+3p_1(1+lp_1)^2}{\left[1-p_1(1+lp_1)^2\right]^3} - \frac{4p_1(1+lp_1)}{\left[1-p_1(1+lp_1)^2\right]^2}, \end{aligned} \tag{4.11}$$

which in relation with Eq.(4.10) shows that

$$|C''(z)| \leq 2lp_1^2 \frac{1+3p_1(1+lp_1)^2}{\left[1-p_1(1+lp_1)^2\right]^3} + \frac{4p_1(1+lp_1)}{\left[1-p_1(1+lp_1)^2\right]^2}. \tag{4.12}$$

Combining Eqs.(4.4), (4.5), (4.8) and (4.11) we can show that $C'(z) < 0$ if the following inequality is true

$$-1 + \frac{2p_1}{(1-p_1)^2} + 2lp_1^2 \frac{1+3p_1(1+lp_1)^2}{\left[1-p_1(1+lp_1)^2\right]^3} < 0. \tag{4.13}$$

We observe that the expression on the left hand side of Eq.(4.13) is increasing in p_1 and since for $p_1 = 0.1$ we have $l \leq 1.1535$ we get

$$\frac{2p_1}{(1-p_1)^2} + 2lp_1^2 \frac{1+3p_1(1+lp_1)^2}{\left[1-p_1(1+lp_1)^2\right]^3} < 0.3 \tag{4.14}$$

which verifies that $C'(z) < 0$.

Now we try to approximate the zero λ. From Lagrange theorem applied on the interval $[1, \lambda]$ we have $C(\lambda) - C(1) = (\lambda - 1)C'(u)$, with $u \in (1, \lambda) \subset (1, 1 + lp_1)$. Since $C(\lambda) = 0$ we get

$$\lambda - 1 = -\frac{C(1)}{C'(u)} \tag{4.15}$$

and taking $\mu = p_1 - p_2 + p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2$ as in Haiman [1999] we obtain

$$\lambda - (1 + \mu) = -\frac{C(1) + \mu C'(u)}{C'(u)}. \tag{4.16}$$

Applying Lagrange theorem one more time as in Eq.(4.4) and combining with Eq.(4.5), the relation in Eq.(4.16) becomes

$$\lambda - (1 + \mu) = -\frac{C(1) + \mu(-1 + 2p_1 - 3p_2) + \mu(R + aC''(1 + \theta a))}{C'(u)}, \tag{4.17}$$

where $a = u - 1$ and $\theta \in (0, 1)$.
If we denote $T_1 = C(1) + \mu(-1 + 2p_1 - 3p_2)$, then

$$\begin{aligned} T_1 &= (p_1 - p_2 + p_3 - p_4 + p_5 - p_6 + \dots) - \mu + (2p_1 - 3p_2)\mu \\ &= (p_5 - p_6 + \dots) + (p_1 - p_2)(2p_1 - 3p_2)^2 + 2(p_1 - p_2)(p_3 - p_4) \\ &\quad -p_2(p_3 - p_4). \end{aligned} \tag{4.18}$$

Using Eq.(4.1) in Eq.(4.18) we observe that

$$-p_1^3 \le -p_2(p_3 - p_4) \le T_1 \le \sum_{k=2}^{\infty} p_1^{k+1} + 4p_1^3 + 2p_1^3$$

which gives us

$$|C(1) + \mu(-1 + 2p_1 - 3p_2)| \le p_1^3 \left[6 + \frac{1}{1 - p_1}\right]. \tag{4.19}$$

Also we can notice that $\mu \ge 0$ and

$$\begin{aligned} \mu &= (1 - p_2)(p_1 - p_2) + p_3 - p_4 + 2(p_1 - p_2)^2 \\ &\le (1 - p_2)(p_1 - p_2) + p_1(p_1 - p_2) + 2(p_1 - p_2)^2 \\ &= 3(p_1 - p_2)^2 + p_1 - p_2 \le p_1(1 + 3p_1). \end{aligned} \tag{4.20}$$

The last step is to find an upper bound for $|C'(u)|^{-1}$. For this observe that

$$\begin{aligned} |C'(u)|^{-1} &= |1 - (2p_1u - 3p_2u^2 + 4p_3u^3 - 5p_4u^4 + \dots)|^{-1} \\ &\le \left|1 - |2p_1u - 3p_2u^2 + 4p_3u^3 - 5p_4u^4 + \dots|\right|^{-1} \end{aligned} \tag{4.21}$$

where we used the inequality $|1 - x| \ge |1 - |x||$. Taking into account that $u \in (1, \lambda) \subset (1, 1 + lp_1)$ and denoting the expression inside the second absolute value in the denominator of Eq.(4.21) by T_2, we have

$$\begin{aligned} T_2 &= \sum_{k=1}^{\infty} 2k(p_{2k-1} - p_{2k}u)u^{2k-1} - \sum_{k=1}^{\infty} p_{2k}u^{2k} \\ &\le 2(p_1 - p_2)u \sum_{k=1}^{\infty} k(p_1u^2)^{k-1} \le \frac{2p_1u}{(1 - p_1u^2)^2}. \end{aligned} \tag{4.22}$$

In the same way

$$\begin{aligned} T_2 &\geq -\sum_{k=1}^{\infty} p_{2k} u^{2k} - 2lp_1 \sum_{k=1}^{\infty} k p_{2k} u^{2k-1} \\ &\geq -\sum_{k=1}^{\infty} p_1^k u^{2k} - 2lp_1^2 u \sum_{k=1}^{\infty} k (p_1 u^2)^{k-1} \\ &\geq -p_1 u \frac{2lp_1 + u(1-p_1u^2)}{(1-p_1u^2)^2} > -\frac{2p_1 u}{(1-p_1u^2)^2}. \end{aligned} \tag{4.23}$$

Combining Eq.(4.22) with Eq.(4.23) along with $u \leq 1 + lp_1$, we have

$$|T_2| \leq \frac{2p_1(1+lp_1)}{[1-p_1(1+lp_1)^2]^2}. \tag{4.24}$$

Substituting Eq.(4.24) in Eq.(4.21) we obtain the bound

$$|C'(z)|^{-1} \leq \frac{1}{1 - \frac{2p_1(1+lp_1)}{[1-p_1(1+lp_1)^2]^2}}. \tag{4.25}$$

Combining the Eqs.(4.8), (4.12), (4.19), (4.20), (4.25) along with the fact that $|a| \leq lp_1$ in Eq.(4.17), we obtain

$$\begin{aligned} |\lambda - (1+\mu)| &\leq \frac{|C(1) + \mu(-1 + 2p_1 - 3p_2)| + |\mu|(|R| + |a||C''(1+\theta a)|)}{|C'(u)|} \\ &\leq K(p_1) p_1^3 \end{aligned} \tag{4.26}$$

where

$$K(p_1) = \frac{\frac{11-3p_1}{(1-p_1)^2} + 2l(1+3p_1)\frac{2+3lp_1 - p_1(2-lp_1)(1+lp_1)^2}{[1-p_1(1+lp_1)^2]^3}}{1 - \frac{2p_1(1+lp_1)}{[1-p_1(1+lp_1)^2]^2}}. \tag{4.27}$$

To obtain $K(\alpha)$ it is enough to substitute p_1 in the above relation with α with the additional remark that $l = l(\alpha)$. □

4.2. **Proof of Theorem 2.3.** For completeness we will give a detailed proof of the theorem even if we repeat most of its ideas from Haiman [1999]. Remembering that

$$C(z) = \sum_{k=0}^{\infty} c_k z^k = 1 + \sum_{k=1}^{\infty} (-1)^k p_{k-1} z^k \tag{4.28}$$

we define

$$D(z) = \frac{1}{C(z)} = \sum_{k=0}^{\infty} d_k z^k \tag{4.29}$$

which exists since $c_0 = 1$, and from $C(z)D(z) = 1$ we have that $d_0 = 1$ and

$$\sum_{j=0}^{n} d_j c_{n-j} = \sum_{j=0}^{n} (-1)^{n-j} p_{n-j-1} d_j = 0, \quad n \geq 1. \tag{4.30}$$

Now if we define $A_k = \{X_1 \leq x, \dots, X_k \leq x, X_{k+1} > x, \dots, X_n > x\}$, for $k = \overline{0,n}$, we obtain $\mathbb{P}(A_0) = p_n$, $\mathbb{P}(A_n) = q_n$ and

$$\mathbb{P}(A_k) + \mathbb{P}(A_{k-1}) = p_{n-k} q_{k-1}, \quad k \geq 1 \tag{4.31}$$

Summing Eq.(4.31) over k, we deduce that

$$q_n = \sum_{k=0}^{n} (-1)^{n-k} p_{n-k} q_{k-1}, \quad n \geq 0, \quad q_{-1} = q_0 = 1 \tag{4.32}$$

and comparing with Eq.(4.30), after using mathematical induction, we conclude that $d_{n+1} = q_n$ and that

$$D(z) = \sum_{k=0}^{\infty} q_{k-1} z^k, \quad q_{-1} = q_0 = 1. \tag{4.33}$$

Taking λ as in Theorem 2.1, we can write $C(z) = U(z)\left(1 - \frac{z}{\lambda}\right)$ and observe that if we let $U(z) = \sum_{k=0}^{n} u_k z^k$ then

$$\left(\sum_{k=0}^{n} u_k z^k\right)\left(1 - \frac{z}{\lambda}\right) = 1 + \sum_{k=1}^{\infty} (-1)^k p_{k-1} z^k \tag{4.34}$$

which shows that

$$u_n - \frac{u_{n-1}}{\lambda} = (-1)^n p_{n-1},\ u_0 = 1,\ n \geq 1. \tag{4.35}$$

Multiplying Eq.(4.35) with λ^n and summing over n we find

$$u_n = \frac{1 + \sum_{k=1}^{n} (-1)^k p_{k-1} \lambda^k}{\lambda^n}, \quad n \geq 1. \tag{4.36}$$

If we denote with $\frac{1}{U(z)} = \sum_{k=0}^{\infty} t_k z^k = T(z)$ then

$$D(z)\left(1 - \frac{z}{\lambda}\right) = T(z) \tag{4.37}$$

and using the same argument as above we get $t_0 = d_0 = 1$ and

$$t_n = d_n - \frac{d_{n-1}}{\lambda}, \quad n \geq 1 \tag{4.38}$$

so $d_n \lambda^n = t_0 + t_1 \lambda + \cdots + t_n \lambda^n$, that is

$$q_n \lambda^{n+1} = t_0 + t_1 \lambda + \cdots + t_n \lambda^n + t_{n+1} \lambda^{n+1}. \tag{4.39}$$

To obtain the desired result we begin by giving an approximation of u_n:

$$\begin{aligned}
|u_n| &= \frac{\left|1 + \sum_{k=1}^{n} (-1)^k p_{k-1} \lambda^k\right|}{\lambda^n} \overset{C(\lambda)=0}{=} \frac{\left|\sum_{k=n+1}^{\infty} (-1)^k p_{k-1} \lambda^k\right|}{\lambda^n} \\
&\leq \frac{\lambda^{n+1}}{\lambda^n} \left|p_n - p_{n+1}\lambda + p_{n+2}\lambda^2 - p_{n+3}\lambda^3 + \dots\right| \\
&\leq \lambda(|p_n - p_{n+1}\lambda| + |p_{n+2} - p_{n+3}\lambda|\lambda^2 + \dots)
\end{aligned} \tag{4.40}$$

Since $\lambda \in (1, 1 + lp_1) \subset (1, 1 + l\alpha)$ we have

$$p_n - p_{n+1}(1 + l\alpha) \leq p_n - p_{n+1}\lambda \leq p_n - p_{n+1} \tag{4.41}$$

which shows that

$$|p_n - p_{n+1}\lambda| \leq p_n - p_{n+1}(1 - l\alpha). \tag{4.42}$$

If we denote by $h = 1 - l\alpha$ and we use the bound from Eq.(4.42) and the fact that $(p_n)_n$ is decreasing in Eq.(4.40) we obtain

$$\begin{aligned}
\frac{|u_n|}{\lambda} &\leq p_n - p_{n+1}h + (p_{n+2} - p_{n+3}h)\lambda^2 + \dots \\
&= (p_n + p_{n+2}\lambda^2 + p_{n+4}\lambda^4 + \dots) - h(p_{n+1} + p_{n+3}\lambda^2 + p_{n+5}\lambda^4 + \dots) \\
&\leq W - h(p_{n+2} + p_{n+4}\lambda^2 + p_{n+6}\lambda^4 + \dots) \\
&= W - \frac{h}{\lambda^2}(W - p_n) = W\left(1 - \frac{h}{\lambda^2}\right) + \frac{h}{\lambda^2}p_n, \qquad (4.43)
\end{aligned}$$

where by Eq.(4.1)

$$\begin{aligned}
W &= p_n + p_{n+2}\lambda^2 + p_{n+4}\lambda^4 + \dots \\
&\leq p_1^{\left[\frac{n+1}{2}\right]} + p_1^{\left[\frac{n+1}{2}\right]}p_1\lambda^2 + p_1^{\left[\frac{n+1}{2}\right]}p_1^2\lambda^4 + \dots \\
&= p_1^{\left[\frac{n+1}{2}\right]}(1 + p_1\lambda^2 + p_1^2\lambda^4 + \dots) = \frac{p_1^{\left[\frac{n+1}{2}\right]}}{1 - p_1\lambda^2}. \qquad (4.44)
\end{aligned}$$

From Eq.(4.43) and Eq.(4.44) we conclude that

$$\begin{aligned}
\frac{|u_n|}{\lambda} &\leq p_1^{\left[\frac{n+1}{2}\right]}\left[\frac{1}{1 - p_1\lambda^2} + \frac{h}{\lambda^2}\left(1 - \frac{1}{1 - p_1\lambda^2}\right)\right] \\
&= \frac{1 - p_1 h}{1 - p_1\lambda^2}p_1^{\left[\frac{n+1}{2}\right]}. \qquad (4.45)
\end{aligned}$$

Until now we have an approximation for u_n, but we still need one for t_n and to solve this aspect lets write

$$T(z) = \frac{1}{U(z)} = \frac{1}{1 - (1 - U(z))} = \sum_{n\geq 0}(-1)^n(U - 1)^n, \qquad (4.46)$$

which is true since the convergence of $C(z)$ implies $|z| < \frac{1}{\sqrt{p_1}}$ so that

$$\begin{aligned}
|1 - U| &\leq \frac{\lambda(1 - p_1 h)}{1 - p_1\lambda^2}\sum_{n\geq 1}p_1^{\frac{n}{2}}|z|^n \leq \frac{|z|\lambda\sqrt{p_1}\left[1 - p_1(1 - lp_1)\right]}{(1 - p_1\lambda^2)(1 - |z|\sqrt{p_1})} \\
&\leq \frac{\sqrt{p_1}(1 + lp_1^2)(1 + lp_1)^2}{\left[1 - p_1(1 + lp_1)^2\right]\left[1 - \sqrt{p_1}(1 + lp_1)\right]} < 0.8. \qquad (4.47)
\end{aligned}$$

Since $u_0 = 1$ we have $U - 1 = \sum_{n\geq 1}u_n z^n$ and

$$(U - 1)^k = \sum_{l\geq 1}\sum_{\substack{i_1+\dots+i_k=l \\ i_j\geq 1, j=\overline{1,k}}} u_{i_1}\dots u_{i_k}z^l. \qquad (4.48)$$

Combining Eq.(4.46) with Eq.(4.48) we get

$$\sum_{n\geq 0}t_n z^n = \sum_{k=0}^{\infty}(-1)^k\sum_{l=1}^{\infty}b_{l,k}z^l, \qquad (4.49)$$

where

$$b_{l,k} = \sum_{\substack{i_1+\dots+i_k=l \\ i_j\geq 1, j=\overline{1,k}}} u_{i_1}\dots u_{i_k}. \qquad (4.50)$$

From Eq.(4.49) we get that $t_0 = 1$ and

$$t_n = \sum_{k=1}^{n} (-1)^k \sum_{\substack{i_1+\cdots+i_k=n \\ i_j \geq 1, j=\overline{1,k}}} u_{i_1} \ldots u_{i_k}, \quad k \geq 1. \tag{4.51}$$

Notice that from Eq.(4.45) we can write

$$|b_{n,k}| \leq \delta^k \sum_{\substack{i_1+\cdots+i_k=n \\ i_j \geq 1, j=\overline{1,k}}} p_1^{\left[\frac{i_1+1}{2}\right]} \ldots p_1^{\left[\frac{i_k+1}{2}\right]}, \tag{4.52}$$

where $\delta = \frac{\lambda(1-p_1 h)}{1-p_1\lambda^2}$. Since, by induction it is easy to verify that

$$\left[\frac{i_1+1}{2}\right] + \cdots + \left[\frac{i_k+1}{2}\right] \geq \left[\frac{i_1+\cdots+i_k+1}{2}\right] = \left[\frac{n+1}{2}\right], \tag{4.53}$$

and the number of terms in the sum of Eq.(4.52) is equal with the number of different positive integers solutions of the equation $i_1 + \cdots + i_k = n$, which is given by $\binom{n-1}{k-1}$, we deduce

$$|b_{n,k}| \leq p_1^{\left[\frac{n+1}{2}\right]} \binom{n-1}{k-1} \delta^k. \tag{4.54}$$

Now from Eq.(4.51) and Eq.(4.54) we have

$$-p_1^{\left[\frac{n+1}{2}\right]} \delta \sum_{k=0}^{\left[\frac{n-1}{2}\right]} \binom{n-1}{2k} \delta^{2k} \leq t_n \leq p_1^{\left[\frac{n+1}{2}\right]} \delta \sum_{k=0}^{\left[\frac{n}{2}\right]-1} \binom{n-1}{2k+1} \delta^{2k+1}, \tag{4.55}$$

so that

$$|t_n| \leq \frac{\delta}{2} p_1^{\left[\frac{n+1}{2}\right]} \left[(1+\delta)^{n-1} + (1-\delta)^{n-1}\right]. \tag{4.56}$$

We see from Eq.(4.39) and Eq.(4.56) that the difference

$$\begin{aligned} |q_n\lambda^n - q_3\lambda^3| &= \left| \frac{\sum_{s=0}^{n+1} t_s\lambda^s - \sum_{s=0}^{4} t_s\lambda^s}{\lambda} \right| = \left| \sum_{s=5}^{n+1} t_s \lambda^{s-1} \right| \\ &\leq \frac{\delta}{2} \sum_{s=5}^{\infty} p_1^{\left[\frac{s+1}{2}\right]} \left[(1+\delta)^{s-1} + (1-\delta)^{s-1}\right] \lambda^{s-1} \end{aligned} \tag{4.57}$$

If we denote by $\sigma_1 = (1+\delta)\lambda$, $\sigma_2 = (1-\delta)\lambda$ and by V the bound in Eq.(4.57), it is not hard to see that

$$V = \frac{\delta p_1^3}{2} \left[\frac{\sigma_1^4(1+\sigma_1)}{1-p_1\sigma_1^2} + \frac{\sigma_2^4(1+\sigma_2)}{1-p_1\sigma_2^2} \right]. \tag{4.58}$$

Recalling that $h = 1 - lp_1$, $\lambda \in (1, 1+lp_1)$ and $p_1 \leq 0.1$ we observe that σ_2 is bounded by

$$-\frac{lp_1[1+2p_1(1+lp_1)]}{1-p_1(1+lp_1)^2} \leq \sigma_2 \leq -\frac{lp_1^2(1+lp_1)}{1-p_1},$$

which gives $|\sigma_2| < 0.5$ and $\frac{\delta\sigma_2^4(1+\sigma_2)}{2(1-p_1\sigma_2^2)} < 0.1$. Substituting the last relation in Eq.(4.58) we can rewrite the bound in Eq.(4.57) as

$$|q_n\lambda^n - q_3\lambda^3| \leq E(p_1)p_1^3 \tag{4.59}$$

where, if we denote by $\eta = 1 + lp_1$,

$$E(p_1) = 0.1 + \frac{\eta^5\left[1+(1-2p_1)\eta\right]^4\left[1+p_1(\eta-2)\right]\left[1+\eta+(1-3p_1)\eta^2\right]}{2(1-p_1\eta^2)^4\left[(1-p_1\eta^2)^2 - p_1\eta^2(1+\eta-2p_1\eta)^2\right]}. \tag{4.60}$$

To obtain $E(\alpha)$ it is enough to make in Eq.(4.60) the substitutions: $p_1 \to \alpha$, $l \to l(\alpha)$ and $\eta \to 1 + l\alpha$. The following lemma gives an approximation for $q_3\lambda^3$.

Lemma 4.1. *If $p_1 \leq \alpha$ then*

$$q_3\lambda^3 = 1 - p_2 + 2p_3 - 3p_4 + p_1^2 + 6p_2^2 - 6p_1p_2 + \mathcal{O}(L(\alpha)p_1^3) \tag{4.61}$$

where $\mathcal{O}(x)$ is a function such that $|\mathcal{O}(x)| \leq |x|$ and $L(\alpha)$ is an expression depending on α and given in Eq.(4.77).

From Eq.(4.57) , Eq.(4.60) and Lemma 4.1 we conclude that

$$|q_n\lambda^n - (1 - p_2 + 2p_3 - 3p_4 + p_1^2 + 6p_2^2 - 6p_1p_2)| \leq \Gamma(\alpha)p_1^3, \tag{4.62}$$

with $\Gamma(\alpha) = L(\alpha) + E(\alpha)$ and this ends the proof of the theorem. □

4.2.1. *Proof of Lemma 4.1.* From Eq.(2.1) of Theorem 2.1 we can write

$$\lambda = 1 + p_1 - p_2 + p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2 + \mathcal{O}(K(\alpha)p_1^3) \tag{4.63}$$

and raising to the third power we get

$$\begin{aligned}\lambda^3 \quad = \quad & (1+\zeta_1+\zeta_2)^3 + 3(1+\zeta_1+\zeta_2)^2\mathcal{O}(K(\alpha)p_1^3) + \mathcal{O}(K^3(\alpha)\alpha^6p_1^3) + \\ & +3(1+\zeta_1+\zeta_2)\mathcal{O}(K^2(\alpha)\alpha^3p_1^3),\end{aligned} \tag{4.64}$$

where we have used the following notations

$$\zeta_1 = p_1 - p_2, \tag{4.65}$$

$$\zeta_2 = p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2. \tag{4.66}$$

Since we can easily see from Eq.(4.65) and Eq.(4.66) that $\zeta_1 = \mathcal{O}(p_1)$ and $\zeta_2 = \mathcal{O}(3p_1^2)$, we deduce that

$$\lambda^3 - (1+\zeta_1+\zeta_2)^3 = \mathcal{O}(S(\alpha)p_1^3) \tag{4.67}$$

with $S(\alpha)$ given bellow by

$$S(\alpha) = 3(1+\alpha+3\alpha^2)^2K(\alpha) + 3\alpha^3(1+\alpha+3\alpha^2)K^2(\alpha) + \alpha^6K^3(\alpha). \tag{4.68}$$

Now we observe that by expanding $(1+\zeta_1+\zeta_2)^3$ we have

$$\begin{aligned}(1+\zeta_1+\zeta_2)^3 \quad = \quad & 1 + 3(\zeta_1+\zeta_2+\zeta_1^2) + 6\zeta_1\zeta_2 + 3\zeta_2^2 + 3\zeta_1\zeta_2^2 + 3\zeta_1^2\zeta_2 \\ & +\zeta_1^3 + \zeta_2^3 \\ = \quad & 1 + 3(\zeta_1+\zeta_2+\zeta_1^2) + \mathcal{O}(18p_1^3) + \mathcal{O}(27\alpha p_1^3) + \mathcal{O}(27\alpha^2p_1^3) \\ & +\mathcal{O}(9\alpha p_1^3) + \mathcal{O}(p_1^3) + \mathcal{O}(27\alpha^3p_1^3)\end{aligned} \tag{4.69}$$

which along with Eq.(4.67) and Eq.(4.68) gives the following expression for

$$\lambda^3 = 1 + 3(\zeta_1+\zeta_2+\zeta_1^2) + \mathcal{O}(P(\alpha)p_1^3), \tag{4.70}$$

where $P(\alpha)$ can be computed by the formula

$$\begin{aligned} P(\alpha) &= 3K(\alpha)(1+\alpha+3\alpha^2)[1+\alpha+3\alpha^2+K(\alpha)\alpha^3]+\alpha^6K^3(\alpha) \\ &\quad +9\alpha(4+3\alpha+3\alpha^2)+19. \end{aligned} \tag{4.71}$$

It is easy to see from Eq.(4.32) that

$$q_3 = 1-p_1-2(p_1-p_2)+p_1^2-p_3 = \mathcal{O}((1-p_1)^2) \tag{4.72}$$

and combined with Eq.(4.70) gives

$$q_3\lambda^3 = q_3[1+3(\zeta_1+\zeta_2+\zeta_1^2)] + \mathcal{O}(P(\alpha)(1-p_1)^2p_1^3). \tag{4.73}$$

The last step in our proof is to find an approximation for the first term on the right in Eq.(4.73). If we write

$$q_3[1+3(\zeta_1+\zeta_2+\zeta_1^2)] = 1-p_2+2p_3-3p_4+p_1^2+6p_2^2-6p_1p_2+H, \tag{4.74}$$

then H checks the relations

$$\begin{aligned} H &= 3(p_1-p_2)\left\{(p_1^2-p_3)[3(p_1-p_2)+1-p_2]+p_2^2-9(p_1-p_2)^2\right\} \\ &\quad -3(p_3-p_4)[p_1+2(p_1-p_2)-(p_1^2-p_3)] \\ &= \mathcal{O}(36p_1^3). \end{aligned} \tag{4.75}$$

Finally, combining Eq.(4.73), Eq.(4.74) and Eq.(4.75) we get

$$q_3\lambda^3 = 1-p_2+2p_3-3p_4+p_1^2+6p_2^2-6p_1p_2+\mathcal{O}(L(\alpha)p_1^3) \tag{4.76}$$

where we used the notation

$$\begin{aligned} L(\alpha) &= 36+(1-p_1)^2P(\alpha) \\ &< 3K(\alpha)(1+\alpha+3\alpha^2)[1+\alpha+3\alpha^2+K(\alpha)\alpha^3]+\alpha^6K^3(\alpha) \\ &\quad +9\alpha(4+3\alpha+3\alpha^2)+55. \quad \square \end{aligned} \tag{4.77}$$

4.3. **Proof of Corollary 2.2 and Corollary 2.4.** To prove the relation in Corollary 2.2 we see that

$$1+p_1-p_2+2(p_1-p_2)^2 = \mu - [p_3-p_4-p_2(p_1-p_2)] \tag{4.78}$$

and since

$$\begin{aligned} p_3-p_4 &= \mathbb{P}(X_1>x, X_2>x, X_3>x, X_4\le x) \\ &\le p_1\mathbb{P}(X_1>x, X_2\le x) = p1(p_1-p_2), \end{aligned} \tag{4.79}$$

we get

$$|p_3-p_4-p_2(p_1-p_2)| \le p_3-p_4+p_2(p_1-p_2) \le p_1^2. \tag{4.80}$$

Combining these relations with Eq.(2.1) from Theorem 2.1, we obtain

$$|\lambda-(1+p_1-p_2+2(p_1-p_2)^2)| \le p_1^2(1+K(\alpha)p_1) \le (1+\alpha K(\alpha))p_1^2 \square \tag{4.81}$$

To prove Corollary 2.4 notice that

$$-3p_1^2 \le p_1^2+2p_3-3p_4+6p_2^2-6p_1p_2 \le 3p_1^2 \tag{4.82}$$

and using Eq.(2.4) from Theorem 2.3, we get

$$\begin{aligned} |q_n\lambda^n-(1-p_2)| &\le \Gamma(\alpha)p_1^3+|p_1^2+2p_3-3p_4+6p_2^2-6p_1p_2| \\ &\le \alpha\Gamma(\alpha)p_1^2+3p_1^2 = (3+\alpha\Gamma(\alpha))p_1^2. \quad \square \end{aligned} \tag{4.83}$$

4.4. **Proof of Theorem 2.5 and Theorem 2.6.** Denoting by

$$\begin{aligned}\mu_1 &= 1 - p_2 + 2p_3 - 3p_4 + p_1^2 + 6p_2^2 - 6p_1p_2,\\ \mu_2 &= 1 + p_1 - p_2 + p_3 - p_4 + 2p_1^2 + 3p_2^2 - 5p_1p_2\end{aligned}$$

we observe that

$$0 \le \frac{\mu_1}{\mu_2} < 1 \tag{4.84}$$

and that

$$\mu_2 = 1 + \underbrace{(p_1 - p_2)(1 - p_2) + p_3 - p_4 + 2(p_1 - p_2)^2}_{\ge 0} \ge 1. \tag{4.85}$$

With the help of Eq.(2.1) and Eq.(2.4) we get

$$\begin{aligned}\left|q_n - \frac{\mu_1}{\mu_2^n}\right| &\le \left|q_n - \frac{\mu_1}{\lambda^n}\right| + \left|\frac{\mu_1}{\lambda^n} - \frac{\mu_1}{\mu_2^n}\right|\\ &\le \Gamma(\alpha)p_1^3 + |\mu_1|\left|\frac{1}{\lambda} - \frac{1}{\mu_2}\right|\left|\frac{1}{\lambda^{n-1}} + \cdots + \underbrace{\frac{1}{\lambda^{n-j}\mu_2^j}}_{\le 1} + \cdots + \frac{1}{\mu_2^{n-1}}\right|\\ &\le [\Gamma(\alpha) + nK(\alpha)]\, p_1^3. \end{aligned}\tag{4.86}$$

To express μ_1 and μ_2 in terms of q's we have to observe first that

$$\begin{aligned}p_1 &= 1 - q_1\\ p_2 &= 1 - 2q_1 + q_2\\ p_3 &= 1 - 3q_1 + 2q_2 + q_1^2 - q_3\\ p_4 &= 1 - 4q_1 + 3q_2 - 2q_1q_2 + 3q_1^2 - 2q_3 + q_4\end{aligned}$$

and after the proper substitutions, we get

$$\mu_1 = 1 + q_1 - q_2 + q_3 - q_4 + 2q_1^2 + 3q_2^2 - 5q_1q_2 \tag{4.87}$$

$$\mu_2 = 6(q_1 - q_2)^2 + 4q_3 - 3q_4 \tag{4.88}$$

To finish the proof of Theorem 2.5 it is enough to use the above relations in Eq.(4.86) to obtain

$$\left|q_n - \frac{6(q_1 - q_2)^2 + 4q_3 - 3q_4}{(1 + q_1 - q_2 + q_3 - q_4 + 2q_1^2 + 3q_2^2 - 5q_1q_2)^n}\right| \le \Delta_1(1 - q_1)^3 \tag{4.89}$$

where

$$\Delta_1 = \Delta_1(\alpha, n) = \Gamma(\alpha) + nK(\alpha).\Box \tag{4.90}$$

For the proof of Theorem 2.6 we use the same approach as in Theorem 2.5. With the help of Eq.(2.3) and Eq.(2.7) we get

$$\begin{aligned}\left|q_n - \frac{\nu_1}{\nu_2^n}\right| &\le \left|q_n - \frac{\nu_1}{\lambda^n}\right| + \left|\frac{\nu_1}{\lambda^n} - \frac{\nu_1}{\nu_2^n}\right| \le (3 + p_1\Gamma(\alpha))p_1^2 + n\nu_1\frac{|\lambda - \nu_2|}{\lambda\nu_2}\\ &\le [3 + \Gamma(\alpha)p_1 + n(1 + p_1K(\alpha))]\, p_1^2\end{aligned}\tag{4.91}$$

where $\nu_1 = 1 - p_2$ and $\nu_2 = 1 + p_1 - p_2 + 2(p_1 - p_2)^2$. We express ν_1 and ν_2 in terms of q's using the relations for p_1 to p_4 from the proof of Theorem 2.5 and we have

$$\nu_1 = 2q_1 - q_2 \tag{4.92}$$

$$\nu_2 = 1 + q_1 - q_2 + 2(q_1 - q_2)^2, \tag{4.93}$$

which we substitute in Eq.(4.91) to obtain

$$\left| q_n - \frac{2q_1 - q_2}{[1 + q_1 - q_2 + 2(q_1 - q_2)^2]^n} \right| \leq \Delta_2 (1 - q_1)^2 \tag{4.94}$$

where

$$\Delta_2 = \Delta_2(\alpha, n, q_1) = 3 + \Gamma(\alpha)(1 - q_1) + n\left[1 + K(\alpha)(1 - q_1)\right].\square \tag{4.95}$$

References

Fu, J.: Distribution of the scan statistic for a sequence of bistate trials. *J. Appl. Prob.* **38** (2001), 908–916.

Glaz, J., Naus, J., Wallenstein, S.: Scan statistic. *Springer* (2001).

Glaz, J., Pozdnyakov, V., Wallenstein, S.: Scan statistic: Methods and Applications. *Birkhauser* (2009).

Haiman, G.: First passage time for some stationary sequence. *Stochastic Processes and their Applications* **80** (1999), 231–248.

Haiman, G., Kiki, M., Puri, M.: Extremes of Markov sequences. *J. Stat Plan Infer* **45** (1995), 185–201.

Haiman, G.: Estimating the distribution of scan statistics with high precision. *Extremes* **3** (2000), 349–361.

Haiman, G., Preda, C.: A new method for estimating the distribution of scan statistics for a two-dimensional Poisson process. *Methodology and Computing in Applied Probability* **4** (2002), 393–407.

Haiman, G., Preda, C.: Estimation for the distribution of two-dimensional scan statistics. *Methodology and Computing in Applied Probability* **8** (2006), 373–381.

Haiman, G.: Estimating the distribution of one-dimensional discrete scan statistics viewed as extremes of 1-dependent stationary sequences. *J. Stat Plan Infer* **137** (2007), 821–828.

Laboratoire de Mathémathiques Paul Painlevé, UMR 8524, Univerité de Sciences et Technologies de Lille 1, France

INRIA Nord Europe/Modal, France

National Institute of R&D for Biological Sciences, Bucharest, Romania
E-mail address: alexandru.amarioarei@inria.fr